Estimation of Data Uncertainty Adjustment Parameters for Multivariate Earth Rotation Series

Li-yu Sung† and J. Alan Steppe
Jet Propulsion Laboratory
California Institute of Technology, Pasadena

Abstract

We have developed a maximum likelihood method to estimate a set of data uncertainty adjustment parameters, including scaling factors and additive variances and covariances, for multivariate Earth rotation series. The necessary condition for maximizing the likelihood function results in a set of nonlinear equations in the unknowns. Partial derivatives of the likelihood function with respect to the unknown parameters arc derived to facilitate the use of nonlinear optimization algorithms in obtaining a numerical solution. 'I he asymptotic covariance matrix of the estimator is derived to provide uncertainty estimates for the parameters. Finally, an example using the data from the Navy VLBI Network (NAVNET) is presented, showing that both the correlations between components and the variation of uncertainty from point to point arc essential elements of the uncertainty structure of the NAVNET data.

v. 5.2

February 3, 1994

To be submitted to:

Manuscript Geodaetica

† Member of Professional Staff, Sterling Software.

Introduction

Earth rotation is inherently multivariate in nature; the components of interest here are Polar Motion X, Polar Motion Y, and UT1. Measurements of earth rotation typically have errors that are correlated between components. The measurement uncertainty also varies from point to point; measurement services typically provide along with a series of measured values a corresponding series of claimed uncertainties that attempts to describe the point to point variation. When an earth rotation time series is smoothed, or when a group of series arc combined, the best possible accuracy in the result can only be obtained if both the correlations between components and the point to point variations in uncertainty are utilized in choosing the weights assigned to the data.

In preparing a smoothing or combination an analyst typically performs a number of intercomparisons to verify the quality of the data; these intercomparisons commonly show that the claimed uncertainties need to be adjusted. Simple commonly used methods of performing this adjustment often result in adjusted uncertainties that do not exhibit correlations between components or point to point variations, or do not adjust these two properties to correspond to the evidence exhibited by the intercomparisons. A smoothing or combination using such adjusted uncertainties thus cannot have the best possible accuracy. Our primary objective in this paper is to present a method of uncertainty adjustment that can adjust both the correlations between components and the point to point variation to correspond to the empirical evidence.

If the actual measurement error contains a component that is statistically independent of those considered in calculating the claimed uncertainties, the addition of a term called the additive variance to the claimed uncertainty will account for this component. If the level of error in the raw data used by the measurement service has been under or over estimated in calculating the claimed uncertainties, this can be corrected by scaling the claimed uncertainties. These two types of adjustments can be used in a multivariate setting if the additive variance is viewed as a covariance matrix and the scaling is applied as a vector of scale factors for the standard deviations as shown below. By using both of these types of adjustment simultaneously we have a method that can adjust both the correlations between components and the point to point variations in uncertainty. Earth rotation measurement series commonly exhibit systematic differences manifested in relative biases and rates; for example, because of differences in the underlying reference frames. In forming a combination series these biases and rates must be estimated and removed. Therefore we have included estimation of bias and rate parameters in our scheme for estimating uncertainty adjustment parameters.

A Kalman filter for earth rotation has been developed at the Jet Propulsion Laboratory for combining and smoothing various series of earth rotation measurements (Eubanks et al. 1985; Morabito et al. 1988). This filter uses a fully multivariate formulation that allows each measured data point to have an associated full covariance matrix. To take full advantage of the capabilities of this filter, we developed the multivariate data uncertainty adjustment technique described in this paper. The following sections present (1) the algorithm for estimating the uncertainty adjustment parameters, (2) a method for calculating the covariance matrix describing the uncertainty in the estimated parameter vector, (3) a method for testing simple statistical hypotheses concerning the uncertainty adjustment parameters, and (4) an example applying these techniques to real data.

The estimation problem

The estimation problem can be stated as follows: given an Earth rotation measurement series, simultaneously estimate the bias, rate, and the uncertainty adjustment parameters (as expressed by the scaling matrix S and the additive covariance matrix **A** detailed below) with respect to an independent reference series:

$$\mathbf{y}_i = \mathbf{b} + \mathbf{r}\Delta t_i + \mathbf{e}_i, \quad i = 1, \cdots, n \qquad [1]$$

where $\mathbf{y}_i$ is a 3x1 column vector of the difference polar motion and UT1-TAI value at the i'th epoch, **b** is a 3x1 column vector of bias, **r** is a 3x1 column vector of rate, Δt_i is the time deviation from a reference time, and $\mathbf{e}_i$ is a 3x1 column vector of random errors having zero mean and a covariance matrix $\mathbf{C}_i$ given by:

$$\mathbf{C}_i = (\mathbf{S}\mathbf{C}_i^p\mathbf{S}^T + \mathbf{A}) + \mathbf{C}_i^r \qquad [2]$$

$\mathbf{C}_i$ is a sum of the adjusted data covariance matrix and the covariance matrix of the reference series $\mathbf{C}_i^r$. The problem of finding a suitable reference series with errors known to be zero mean and covariance matrix $\mathbf{C}_i^r$ is an important one, but is beyond the scope of this paper.The observed data covariance matrix $\mathbf{C}_i^o$ is adjusted using matrices S and **A** with the following form:

$$\mathbf{S} = \begin{pmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & m_z \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 & \sigma_{xz}^2 \\ \sigma_{xy}^2 & \sigma_y^2 & \sigma_{yz}^2 \\ \sigma_{xz}^2 & \sigma_{yz}^2 & \sigma_z^2 \end{pmatrix} \qquad [3]$$

The product $\mathbf{S}\mathbf{C}_i^o\mathbf{S}^T$ scales the variances and covariances of $\mathbf{C}_i^o$ by the product of the corresponding components in S. Although different ways of adjusting the formal error of the observed data are possible, the model considered here is quite general in practice. Given $\mathbf{y}_i, \Delta t_i, \mathbf{C}_i^o$ and $\mathbf{C}_i^r$, the objective is to solve for **b, r,** and the uncertainty adjustment parameter vector **m** defined as:

$$\mathbf{m} = (m_x \;\; m_y \;\; m_z \;\; \sigma_x^2 \;\; \sigma_y^2 \;\; \sigma_z^2 \;\; \sigma_{xy}^2 \;\; \sigma_{xz}^2 \;\; \sigma_{yz}^2)^T \qquad [4]$$

Maximum Likelihood Solution

1' he estimation problem considered here has the unknown **m** embedded in $\mathbf{C}_i$ and therefore the standard least squares method, in which the covariance matrix $\mathbf{C}_i$ is assumed to be known (to within a scaling factor), will not work, The maximum likelihood method is used instead to find the solution. Assuming that the $\{\mathbf{e}_i\}$ are mutually independent with each following a multivariate normal distribution, the likelihood function $\mathcal{L}$ is given by:

$$\begin{aligned} \mathcal{L}(\mathbf{y}_1, \cdots, \mathbf{y}_n, \mathbf{b}, \mathbf{r}, \mathbf{m}) &= \prod_{i=1}^{n} p(\mathbf{y}_i \mid \mathbf{b}, \mathbf{r}, \mathbf{m}) \\ &= \prod_{i=1}^{n} P(\mathbf{e}_i \mid \mathbf{b}, \mathbf{r}, \mathbf{m}) \\ &= (2\pi)^{-3n/2} \left(\prod_{i=1}^{n} |\mathbf{C}_i|^{-1/2} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{e}_i^T \mathbf{C}_i^{-1} \mathbf{e}_i) \right\} \end{aligned}$$

where P denotes the probability density function and $|\mathbf{C}_i|$ is the determinant of $\mathbf{C}_i$. Note that the unknowns **b** and **r** are embedded in e_i while m is embedded in $\mathbf{C}_i$. Given a set of observations $\{\mathbf{y}_i\}$, the maximum likelihood estimate of **b**, **r,** and **m** is chosen as the one which maximizes the likelihood function $\mathcal{L}$. Since $\mathcal{L}$ can also be written as:

$$\mathcal{L}(\mathbf{b}, \mathbf{r}, \mathbf{m}) = (2\pi)^{-3n/2} \exp(-F(\mathbf{b}, \mathbf{r}, \mathbf{m}))$$

and

$$F(\mathbf{b}, \mathbf{r}, \mathbf{m}) = \frac{1}{2}\sum_{i=1}^{n} \ln|\mathbf{C}_i| + \frac{1}{2}\sum_{i=1}^{n} (\mathbf{e}_i^T \mathbf{C}_i^{-1} \mathbf{e}_i) \qquad [5]$$

the maximum likelihood estimate equivalently minimizes F. It follows from equation [5] that the estimate minimizes a combination of the weighted sum of residual squares together with the overall magnitude (as represented by the determinant of $\mathbf{C}_i$) of the covariance matrices. To clearly see the nonlinear nature of the problem, consider the necessary condition for minimizing F:

$$\begin{vmatrix} \frac{\partial F}{\partial \mathbf{b}} \\ \frac{\partial F}{\partial \mathbf{r}} \\ \frac{\partial F}{\partial \mathbf{m}} \end{vmatrix} = \begin{vmatrix} -\sum_{i=1}^{n} \mathbf{C}_i^{-1}(\mathbf{y}_i - \mathbf{b} - \mathbf{r}\Delta t_i) \\ -\sum_{i=1}^{n} \Delta t_i \mathbf{C}_i^{-1}(\mathbf{y}_i - \mathbf{b} - \mathbf{r}\Delta t_i) \\ \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial \mathbf{m}} \ln|\mathbf{C}_i| + \frac{\partial}{\partial \mathbf{m}}(\mathbf{e}_i^T \mathbf{C}_i^{-1} \mathbf{e}_i)\right) \end{vmatrix} = \mathbf{0} \qquad [6]$$

The above set of equations is clearly nonlinear in the unknown parameters and the solution will require iteration.

For practical application of the covariance adjustment parameters it is important that the adjusted matrix $\mathbf{S}\mathbf{C}^o\mathbf{S}^T + \mathbf{A}$ be a legitimate covariance matrix for new data with an arbitrary covariance matrix C^o as well as for the given matrices $\mathbf{C}_i^o$. Thus $\mathbf{S}\mathbf{C}^o\mathbf{S}^T + A$ must be non- negative definite for arbitrary $\mathbf{C}^o$ and hence A must be non- negative definite. Let A be expressed by the Cholesky decomposition $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ where **L** is a lower triangular matrix:

$$\mathbf{L} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix}$$

then **A is** non--negative definite for any **L.** For simplicity **A is** required here to be strictly positive definite by requiring the diagonal elements of **L** to be strictly positive; thus the **L** corresponding to a particular **A** is unique. In this way, we are assured that σ_x^2, σ_y^2, and σ_z^2 are positive, but in general $\sigma_{xy}^2, \sigma_{xz}^2$, and σ_{yz}^2 could possibly be negative.

Since changing the sign of all three scaling factors m_x, m_y, m_z does not affect the adjusted covariance matrix, it can be assumed without loss of generality that $m_x \geq 0$.

Solution Algorithm

As the function F is a nonlinear function of the unknown parameters, a numerical solution can best be obtained by using nonlinear optimization techniques (e.g. Gill et al., 1981). A package of Fortran subroutines called NPSOL (Gill et al., 1986) is used here to minimize F (given by equation [5]) subject to simple bounds: $m_x \geq 0$, $l_{11}>0$, $l_{22} > 0$, and $l_{33}>0$.

Starting from an initial set of values for b, **r, S,** and **L,** NPSOL uses a sequential quadratic programming algorithm to find the solution (see Gill et al., 1981, 1986 for details). Although NPSOL can approximate the partial derivatives of F with respect to the unknown parameters by finite differences, it is more efficient and reliable if these partial derivatives are given analytically.

Partial Derivatives

The partial derivatives $\partial F/\partial \mathbf{b}$ and $\partial F/\partial \mathbf{r}$ are given in [6]. The partial derivative $\partial F/\partial \mathbf{m}$ is shown in Appendix A to be equal to :

$$\frac{\partial F}{\partial \mathbf{m}} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Gamma_i^T}{\partial \mathbf{m}}\right)(\mathbf{g}_i - \mathbf{f}_i) \qquad [7]$$

where

$$\mathbf{f}_i = (C_{11_i}\ \ C_{22_i}\ \ C_{33_i}\ \ C_{12_i}\ \ C_{13_i}\ \ C_{23_i})^T$$

$$\mathbf{g}_i = (e_{x_i}^2\ \ e_{y_i}^2\ \ e_{z_i}^2\ \ e_{x_i}e_{y_i}\ \ e_{x_i}e_{z_i}\ \ e_{y_i}e_{z_i})^T$$

$$\frac{\partial \Gamma_i^T}{\partial \mathbf{m}} = \left(\frac{\partial \Gamma_{11_i}}{\partial \mathbf{m}} \; \frac{\partial \Gamma_{22_i}}{\partial \mathbf{m}} \; \frac{\partial \Gamma_{33_i}}{\partial \mathbf{m}} \; \frac{\partial (2\Gamma_{12_i})}{\partial \mathbf{m}} \; \frac{\partial (2\Gamma_{13_i})}{\partial \mathbf{m}} \; \frac{\partial (2\Gamma_{23_i})}{\partial \mathbf{m}} \right)$$

and the C_{kj_i}'s are the elements of $\mathbf{C}_i$, the e_{k_i}'s are the elements of $\mathbf{e}_i$ and the Γ_{kj_i}'s are the elements of Γ_i, which is the inverse matrix of $\mathbf{C}_i$. Differentiating $\mathbf{C}_i^{-1}\mathbf{C}_i = 1$ with respect to the h'th element of $\mathbf{m}$ yields:

$$\frac{\partial \mathbf{C}_i^{-1}}{\partial m_h} = -\mathbf{C}_i^{-1} \frac{\partial \mathbf{C}_i}{\partial m_h} \mathbf{C}_i^{-1}$$

which can be used to compute the matrix $\frac{\partial \Gamma_i^T}{\partial \mathbf{m}}$. Partial derivatives with respect to the elements of L, which are needed by NPSOL, can be obtained from the partial derivatives with respect to the elements of A by applying the chain rule, yielding:

$$\frac{\partial F}{\partial L_{jk}} = 2 \sum_m \frac{\partial F}{\partial A_{jm}} L_{mk}$$

Uncertainty of the Parameter Estimates

Let $\hat{\mathbf{x}}$ be the maximum likelihood estimator of x, then it can be shown (e.g. Kendall and Stuart, 1979) that, when the number of data becomes large, $\hat{\mathbf{x}}$ has a distribution which becomes normal, with mean x and a covariance matrix V given by:

$$\mathbf{V}^{-1} = \left\langle \left(\frac{\partial \ln \mathcal{L}}{\partial \mathbf{x}}\right)\left(\frac{\partial \ln \mathcal{L}}{\partial \mathbf{x}}\right)^T \right\rangle = \left\langle \left(\frac{\partial F}{\partial \mathbf{x}}\right)\left(\frac{\partial F}{\partial \mathbf{x}}\right)^T \right\rangle \qquad [8]$$

Substituting the partial derivatives into [8] and carrying out the expectation:

$$\mathbf{V}^{-1} = \begin{bmatrix} \sum_{i=1}^{n} \mathbf{C}_i^{-1} & \sum_{i=1}^{n} \Delta t_i \mathbf{C}_i^{-1} & \mathbf{0} \\ \sum_{i=1}^{n} \Delta t_i \mathbf{C}_i & \sum_{i=1}^{n} \Delta t_i^2 \mathbf{C}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{4} \sum_{i=1}^{n} \left(\frac{\partial \Gamma_i^T}{\partial \mathbf{m}}\right) \mathbf{E}_i \left(\frac{\partial \Gamma_i^T}{\partial \mathbf{m}}\right)^T \end{bmatrix} \qquad [9]$$

where $E_i = \langle(g_i - f_i)(g_i - f_i)^T\rangle$ **is** derived in Appendix B. The zero elements result from the fact that the third moments of e_i, as well as the first, are zero since e_i is assumed to be normally distributed. It follows that $\hat{b}$ and $\hat{r}$ are, asymptotically, statistically independent of m. Note that **V is** a function of the unknown parameter m. To obtain an estimate of **V,** one simply substitutes m for **m** in [9].

Example

As an example of the estimation problem discussed above, we have estimated the biases, rates, and data uncertainty adjustment parameters for a set of NAVNET VLBI data (Eubanks et al., 1993) from 1989/9/1 1 to 1992/12/31 with respect to a smoothed reference series SPACE92 (Gross 1993). A total of 220 measurement epochs are used. The primary reason to use NAVNET data is that the full data covariance matrix is available.

l-he results obtained using NPSOL are shown in Table 1 in the column entitled 'Full solution'. The corresponding value of F is 628.451. We then tested several hypotheses about the error structure of the NAVNET data using the likelihood ratio test described below and the results arc shown in Table 2, where **I** and O are the identity and zero matrices respectively.

For example, the null hypothesis H_0 : S = **1, A** = **O** can be tested against the alternative hypothesis H_a : S ≠ 1, **A** ≠ **O** using the asymptotic distribution of the likelihood ratio $\lambda = P_0/P_f$, where P_0 is the likelihood of the null model and P_f is the likelihood of the full model in which S and **A** are unrestricted. For large n, it can be shown that $-2\ln\lambda$ has asymptotically a χ^2 distribution with degrees of freedom (p) equal to the number of parameters tested in the hypothesis (e.g. Mendenhall et al., 1990). 1 herefore, the rejection region is given by $-2\ln\lambda > \chi^2(p, \alpha)$ where α is the probability that such a χ^2 value will be exceeded if the null hypothesis is correct.

Case 1 tested the hypothesis that no uncertainty adjustments are needed, case 2 tested the hypothesis that all scaling factors are equal to zero (i.e. one can completely ignore the provided data covariance matrices), case 3 tested the hypothesis that no additive variances and covariances are needed, case 4 tested the hypothesis that no additive off--diagonal covariances are needed, case 5 tested the hypothesis that no scalings are needed, case 6 tested the hypothesis that all three scaling factors arc equal. All hypotheses are rejected at **99°/0** confidence level except case 6, which suggests that a single scaling factor can be used to scale the uncertainty in addition to **A**.

Discussion

For comparison, we have determined another solution in which the uncertainty adjustment parameters were estimated one component at a time. In this case the claimed correlations between components are necessarily neglected and the off- diagonal elements of **A** cannot be estimated. The result is shown in Table 1. When the resulting parameter values are used in equation [5] to make a (multivariate) evaluation of F, the value obtained is 649.333, corresponding to a likelihood that is $9x10^{-}$ 10 times that of the full model. I-he small likelihood is due in part to the neglect of the off- diagonal terms of **A,** and in part to the neglect of the off- diagonal elements of $\mathbf{C}_i^{\rho}$ and $\mathbf{C}_i^{r}$ when estimating the parameters. Even though the individual parameter values are not significantly different from the full solution except σ_{xz}^2 , the solution when viewed as a vector of dimension 15 lies outside the **99%** confidence region given by the full solution.

The maximum likelihood method thus provides a means for objectively deciding such questions as (1) whether correlations between components must be accounted for, (2) whether variations of uncertainty between data points must be accounted for, (3) whether uncertainty adjustment by scaling or by use of additive variance is more effective, and (4) whether both **S** and **A** are needed.

The likelihood function $\mathcal{L}$ can have more than one local maximum and which one is found by NPSOL can depend upon the starting value of S and **L.** However, we found that, for the NAVNET example, as long as the starting values of l_{11}, l_{22}, anti l_{33} are not too close to zero (i.e. too near to the lower bounds) and all scaling factors are constrained to be positive (i.e. $\mathbf{S}\mathbf{C}_i^{\rho}\mathbf{S}^T$ should preserve the sign as well as the magnitude of the original correlations) , a variety of starting points all yielded a single solution, which had a larger likelihood than was sometimes obtained if the above conditions were not met. Note that allowing m_y and m_z to be negative (and starting at say, my $= \pm 1$ $m_z = \pm 1$) can be used to test for sign errors in the claimed correlations (we did not find any in the NAVNET example). A good starting point is provided by **b = O, r = O, S = 1,** and **L = I** (in data units).

The maximum likelihood solution does not require that the reduced chisquare of the residuals be one. However, in all the examples we have tried, the optimal solution has had a reduced chisquare that is not statistically significantly different from one. In the NAVNET example, the reduced chisquare of the full solution is 1.04.

The maximum likelihood approach presented here can be readily extended to a higher dimension than three, for example, by including the nutation angles in the adjustment.

Acknowledgements

We thank James R. Jee for pointing out the asymptotic property of maximum likelihood estimators, and Charles L. Lawson and Van Snyder for providing the NPSOL software package used here. The work described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Appendix A: $\partial F/\partial \mathbf{m}$

From equation [5], the partial derivative can be written as:

$$\frac{\partial F}{\partial \mathbf{m}} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial \mathbf{m}}\ln\left|\mathbf{C}_i\right| + \frac{\partial}{\partial \mathbf{m}}\left(\mathbf{e}_i^T\mathbf{C}_i^{-1}\mathbf{e}_i\right)\right) \qquad [10]$$

Let

$$\mathbf{e}_i = \begin{pmatrix} e_{x_i} \\ e_{y_i} \\ e_{z_i} \end{pmatrix}, \quad \mathbf{C}_i = \begin{pmatrix} C_{11_i} & C_{12_i} & C_{13_i} \\ C_{12_i} & C_{22_i} & C_{23_i} \\ C_{13_i} & C_{23_i} & C_{33_i} \end{pmatrix}, \quad \boldsymbol{\Gamma}_i = \mathbf{C}_i^{-1} = \begin{pmatrix} \Gamma_{11_i} & \Gamma_{12_i} & \Gamma_{13_i} \\ \Gamma_{12_i} & \Gamma_{22_i} & \Gamma_{23_i} \\ \Gamma_{13_i} & \Gamma_{23_i} & \Gamma_{33_i} \end{pmatrix}$$

$$\mathbf{f}_i = \left(C_{11_i} \;\; C_{22_i} \;\; C_{33_i} \;\; C_{12_i} \;\; C_{13_i} \;\; C_{23_i}\right)^T$$

$$\mathbf{g}_i = \left(e_{x_i}^2 \;\; e_{y_i}^2 \;\; e_{z_i}^2 \;\; e_{x_i}e_{y_i} \;\; e_{x_i}e_{z_i} \;\; e_{y_i}e_{z_i}\right)^T$$

$$\boldsymbol{\Gamma}_i = (\Gamma_{11_i}\ \Gamma_{22_i}\ \Gamma_{33_i}\ 2\Gamma_{12_i}\ 2\Gamma_{13_i}\ 2\Gamma_{23_i})^T$$

the second term of [10] can be written as:

$$\frac{\partial}{\partial \mathbf{m}}(\mathbf{e}_i^T \mathbf{C}_i^{-1}\mathbf{e}_i) = \frac{\partial}{\partial \mathbf{m}}(\boldsymbol{\Gamma}_i^T \mathbf{g}_i) = \left(\frac{\partial \boldsymbol{\Gamma}_i^T}{\partial \mathbf{m}}\right)\mathbf{g}_i \qquad [11]$$

and the first term of [10] can be written as:

$$\frac{\partial}{\partial \mathbf{m}}\ln|\mathbf{C}_i| = \frac{1}{|\mathbf{C}_i|}\frac{\partial|\mathbf{C}_i|}{\partial \mathbf{m}} = \frac{1}{|\mathbf{C}_i|}\sum_{jk}\frac{\partial|\mathbf{C}_i|}{\partial C_{jk_i}}\frac{\partial C_{jk_i}}{\partial \mathbf{m}} = \sum_{jk}\Gamma_{kj_i}\frac{\partial C_{jk_i}}{\partial \mathbf{m}}$$

$$= \mathrm{tr}\left(\boldsymbol{\Gamma}_i\frac{\partial \mathbf{C}_i}{\partial \mathbf{m}}\right) = \left(\frac{\partial \mathbf{f}_i^T}{\partial \mathbf{m}}\right)\boldsymbol{\Gamma}_i = -\left(\frac{\partial \boldsymbol{\Gamma}_i^T}{\partial \mathbf{m}}\right)\mathbf{f}_i \qquad [12]$$

The last equality of [12] follows from the fact that $\mathbf{f}_i^T\boldsymbol{\Gamma}_i = 3$. Equation [7] is obtained by combining equations [11] and [12] with [10].

Appendix B: Calculation of E_i

For simplicity, the subscript i is dropped in the following derivation. Recall the definition of E:

$$\mathbf{E} = \langle(\mathbf{g}-\mathbf{f})(\mathbf{g}-\mathbf{f})^T\rangle = \langle\mathbf{g}\,\mathbf{g}^T\rangle - \mathbf{f}\,\mathbf{f}^T$$

where the expectation typically involves $\langle e_k\, e_l\, e_m\, e_n\rangle$, which can be evaluated using the characteristic function of a zero mean multivariate normal distribution (e.g. Chatfield and Collins, 1980)

$$\Phi_{\mathbf{e}}(\mathbf{t}) = \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{q} C_{ij} t_i t_j\right\}$$

with q being the number of components in each measurement vector (3 here). The result is:

$$\langle e_k\, e_l\, e_m\, e_n\rangle = \frac{1}{i^4}\frac{\partial^4\Phi_{\mathbf{e}}(\mathbf{t})}{\partial t_k\partial t_l\partial t_m\partial t_n}\bigg|_{\mathbf{t}=\mathbf{0}} = C_{kl}C_{mn} + C_{km}C_{ln} + C_{kn}C_{lm}$$

Now, the elements of E can "be written as:

$$E_{ij} = \langle g_i g_j \rangle - f_i f_j = \langle e_k e_l e_m e_n \rangle - C_{kl}C_{mn} = C_{km}C_{ln} + C_{kn}C_{lm}$$

where the indices (k,l) and (m,n) correspond respectively to the i'th and j'th components of both g and **f**. Note that the above derivation is valid for an arbitrary dimension q.

References

Chatfield, C. and A.J. Collins, Introduction to Multivariate Analysis, Chapman and Hall, 1980.

Eubanks, T. M., B.A. Archinal, M.S. Carter, F.J. Josties, D.N. Matsakis, and D. C). McCarthy, Earth Orientation Results From the U.S. Naval Observatory VLBI Program, in NEOS Annual Report for 1992, 1993.

Eubanks, T. M., J.A. Steppe, and M.A. Spieth, The Accuracy of Radio Interferometric Measurements of Earth Rotation, TDA Progress Report 42-80, October- December 1984, Jet Propulsion Laboratory, 1985.

Gill, P. E., W. Murray, anti M.H. Wright, Practical Optimization, Academic Press, 1981.

Gill, P. E., W. Murray, M.A. Saunders and M. F{. Wright, User's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming, l-ethnical Report SOL 86-2, Department of Operations Research, Stanford University, 1986.

Gross, R.S. , A Combination of Earth Orientation Data: SPACE92, International Earth Rotation Service l-ethnical Note 14, Observatoire de Paris, 1993.

Kendall, M. and A. Stuart, The Advanced Theory of Statistics, Vol 2. Inference & Relationship, 4th Ed., Charles Griffin & Company Ltd., England, 1979.

Mendenhall, W., D.D. Wackerly, and R.L. Scheaffer, Mathematical Statistics with Applications, 4th Ed., PWS- KEN-I Publishing Company, Boston, 1990.

Morabito, D. D., T.M. Eubanks, and J.A. Steppe, Kalman Filtering of Earth Orientation Changes, in The Earth's Rotation and Reference Frames for Geodesy and Geodynamics (Proceedings of the 128th Symposium of the International Astronomical Union, Coolfont, West Virginia, 20- 24 October, 1986), pp. 257-267, Kluwer Acad. Publ., Dordrecht, Holland, 1988.

Table 1: Parameter Estimates

Quantity	Full solution	Component by component solution	Unit
b_x	-0.40±0.05	-0.39±0.06	mas
b_y	0.97±0.05	0.98±0.05	mas
b_z	1.38±0.05	1.42±0.05	mas
r_x	0.24±0.04	0.23±0.04	mas/yr
r_y	-0.06±0.04	-0.06±0.04	mas/yr
r_z	0.03±0.03	0.00±0.03	mas/yr
m_x	1.52±0.16	1.49±0.20	
m_y	1.24±0.15	1.33±0.18	
m_z	1.32±0.15	1.23±0.19	
σ_x^2	0.09±0.03	0.10±0.04	mas^2
σ_y^2	0.12±0.03	0.10±0.03	mas^2
σ_z^2	0.08±0.03	0.08±0.03	mas^2
σ_{xy}^2	0.00±0.02		mas^2
σ_{xz}^2	-0.07±0.01		mas^2
σ_{yz}^2	-0.03±0.02		mas^2
F	628.451	649.333	

Table 2: Case studies

Case	H_0	p	$-2\ln\lambda$	$\chi^2(p,0.01)$
1	**A = 0,S = I**	9	271.2	21.7
2	**S = 0** “	3	178.6	11,3
3	**A = O** “’	6	42.5	16.8
4	**A** diagonal	3	26.2	11.3
5	**S = I**	3	12.8	11.3
6	S = ml	2	2,9	9.21

p = number of parameters tested

λ = likelihood ratio: $\dfrac{\text{likelihood of the null model}}{\text{likelihood of the full model}}$

$\chi^2(p,0.01)$ = right- tail critical number for p degrees of freedom and $\alpha = 0.01$